

⑨ 日本国特許庁 (JP)  
⑩ 公開特許公報 (A)

⑪ 特許出願公開  
昭57-163299

⑫ Int. Cl.<sup>3</sup>  
G 10 L 1/00

識別記号

庁内整理番号  
7350-5D

⑬ 公開 昭和57年(1982)10月7日

発明の数 1  
審査請求 未請求

(全 5 頁)

⑭ 音声合成装置

⑯ 特 願 昭56-47143  
⑰ 出 願 昭56(1981)4月1日  
⑱ 発 明 者 森戸誠  
東京都港区虎ノ門1丁目7番12

号沖電気工業株式会社内  
⑲ 出 願 人 沖電気工業株式会社  
東京都港区虎ノ門1丁目7番12  
号  
⑳ 代 理 人 弁理士 鈴木敏明

M.

明 細 書

1. 発明の名称

音声合成装置

2. 特許請求の範囲

無声音素片を再生し得る多数の単位を記憶させておき、合成すべき音声に対応して無声音素片を選択し、合成音を出力する音声合成装置において、前記各単位は自然無声音素片と当該自然無声音素片の時間軸を逆にしたものとの相加平均の $\frac{1}{2}$ 時間長を表現するものであり、任意の前記単位における一連のデータを順方向及び逆方向に連続して各1回ずつ読み出して無声音素片を再生する手段を備えていることを特徴とする音声合成装置。

3. 発明の詳細な説明

本発明は、記憶装置から音声波形の波形領域の情報を読み出し、音声を合成する音声合成装置に関する。

一般に音声合成装置は出力しようとするすべての言葉を文節単位、単語単位あるいは音いん単位に分類し、記憶装置に格納し、出力すべき言葉

の情報にしたがって前記の分類した音声をつなぎあわせ、1つの文章として出力している。

この種の音声合成装置において、言葉をどの程度の単位で分類し記憶しておくかということは、合成装置の規模あるいは音質を決定する重要な問題となる。すなわち、文節単位、あるいは文章に近い単位で分類し記憶すると音と音をつなぎあわせる操作が少なくなり、加えて文節の抑揚も含めて言葉を記憶するため音質はきわめて良好となる。反面、出力する言葉のすべてを記憶しておかなければならないため記憶領域が莫大になる。第1図にこの方式にしたがった記憶方法を示す。

次に、音節単位で分類し記憶する方法を考える。日本語は100個程度の音節を持っているためこれだけの音節を記憶装置に格納しておけば日本語すべてが合成できる。しかし、反面言葉の抑揚・強弱などの制御が複雑となる。勿論、すべての抑揚、すべての強弱について100音節それぞれ記憶することによって多少は軽減されるものの音節と音節のつなぎ方が複雑で自然性をもった合成音

声出力を得ることはきわめて困難である。第2図にこの方式にしたがう記憶方法を示す。

したがって、現存では前記両者の中間的な方法が多くとられている。すなわち、合成音声の固定部分についてはできるだけ長い文章単位で記憶装置に格納し、可変部分の中で選択的な部分（いくつかの言葉の中から選ぶ場合、たとえば数字、単位など）については単語単位で記憶装置に格納し、可変部分の中でも非選択的な部分（たとえば名前、地名などの固有名詞など）については音節単位で記憶装置に格納する方法である。第3図にこの方式にしたがう記憶方法を示す。

これらの方法においては記憶装置の記憶領域を軽減するためには重複する単語等を1つにする方法がとられる。すなわち、音声の単位を1つのブロックと考えた場合、ブロックの個数を減らすことによって記憶装置の記憶領域を小さくする方法である。

一方、前記の方法とは見方を異にする記憶領域の軽減方法がある。それは音声の単位となる1つ

の音声「シ」の波形を示す。第5図(b)に示されるように無声音においては有声音のような規則的な性質はなくピッチも存在しない。そのため無声音においては適当なフレームごとに波形そのものを記憶しておくしか手段がなく圧縮効果は期待できない。しかし、無声音の中にはまざつ音（f, s）などのように持続時間が長い音もあるため無声音に対する圧縮効果も音声の記憶領域の軽減には重要な問題となる。

本発明は無声音時の音声素片を時間対称化させ、無声音の音声素片を格納する音声合成装置の記憶装置の記憶領域を軽減するようにしたもので、以下詳細に説明する。

第6図に、入力音声が無声音、有声音の判定を行って無声音に関しては波形の対称化を行い、音声合成装置の記憶装置に格納するまでの図を示す。以下、第6図について説明する。

入力される音声はある一定の標本化周期でサンプリングされ、ある一定の分析区間Tにおいて、判定回路1で有声音、無声音の判定を行う。その

ブロックに要する記憶領域を軽減する方法である。一般に記憶装置にたくわえられる音声は125  $\mu$ 秒ごとにA/D変換器によって12ビットのパターンに変換される。したがって、1秒の音声には96Kビットの記憶領域が必要となる。この領域を軽減する方法としては多くの研究がなされており、代表的な方法にLog-PCM, DPCM, ADM, ADPCM, PARCOR, LSPがある。これらの方法は音声の性質を利用してあり、これから述べる方法もその一部といえよう。

それは音声の周期性を利用した方法である。第4図に音声波形の波形図を示す。第4図に示されるように音声はきまった周期で同一波形に近い波形がくりかえしており、その周期はピッチと呼ばれる。そこでピッチ単位で隣接する波形がよく似ていることに着目し、同一波形を2回、3回と繰り返すことによって記憶領域を軽減する方法がある。しかし、このように音声波形がピッチごとの規則的な性質を示すのはその音声が有声音の場合だけである。第5図(a)に有声音の「ア」、第5図(b)に無

結果、無声音と判定された波形をf(i)とする。ただし、分析区間Tは長くても声道の伝達関数がほぼ一定である30ms程度とする。これは音声のスペクトラム形状が変化していないと考えられる一般的な時間区間であり、仮に実際の無声音が30msec以上、例えば90msec継続していたとしても無声音の一つの区間は30msec程度とする。もし、90msecの区間無声音が継続していた場合には30msecの無声音を一つの単位としてそれが3つ続いていると考える。ここで一つの無声音の区間を便宜上30msecと限定し、その区間での入力された無声音波形f(i)を考える。第7図(a)に無声音波形の例を示す。

次に、波形対称化回路2において、無声音の波形の対称化処理を行う。無声音の波形対称化処理について次に説明する。

分析区間T（例えば30ms）で無声音と判定された波形f(i), ( $0 \leq i \leq 30ms$ )に対するこの分析区間Tでの無声音対称波形g(i)は次式(ii)の如き演算を行って求められる。

$$g(i) = \frac{1}{2} \{ f(i) + f(30-i) \} \quad 0 \leq i \leq 30 \text{ ms} \quad \dots\dots (1)$$

ここで  $f(30-i)$  は前記無声音波形  $f(i)$  の時間軸を逆にした時の波形であり、第7図(h)に示す。又無声音対称波形  $g(i)$  の波形図を第7図(c)に示す。前記(1)式並びに第7図(c)を参照すればわかるとおり、無声音対称波形  $g(i)$  は分析区間  $T$  の中央である  $i = 15$  に対して対称な波形となる。

次に、対称化された無声音対称波形  $g(i)$  の分析区間  $T$  中の半分の区間  $0 \leq i \leq 15 \text{ ms}$  のサンプル点に対応する波音値データを音声合成装置の記憶装置3に格納する。尚、記憶装置3には音声素片の長さ(データの数)及びくり返し回数も記憶される。

第8図は、本発明の音声合成装置の記憶装置の記憶構造を示した図である。第8図(a)に示した構造において、大ブロックはその各々が単語等の語長に相当する音声ファイルを格納するもので、1つの大ブロックは第8図(b)に示す複数の小ブロックからなっている。小ブロックはその1つ1つが1つの有声音又は無声音の音声素片に対応してお

り、無声音素片に対応する1つの小ブロックは第8図(c)に示すような構造で、音声素片の長さ(データの数)、くりかえし回数、並びに前述の波形対称化回路で抽出した無声音対称化波形  $g(i)$  の分析区間  $T$  の半分の区間長における各サンプル点の波音値データ(データ  $\#1$  - データ  $\#m$ ) が格納される。尚、大ブロックの先頭アドレスやそれに含まれる小ブロックの個数は第8図(d)に示す如く、記憶装置3の先頭部にあるテーブルに格納しておく。

第9図は、本発明による音声合成装置を示したもので、記憶装置3、コントローライ、アドレスカウンタ5、D/A変換器6、ローパスフィルタ7、音声出力部8からなる。

まず、コントローライに対する外部からの音声ファイル読み出し指令により起動がかかると、コントローライは記憶装置3のテーブルを参照し、読み出すべき音声ファイルの大ブロックのアドレス、小ブロックの個数等を読み出す。この読み出した大ブロックのアドレスをアップ・ダウンカウ

ンタであるアドレスカウンタ5に設定し、該当の大ブロックに含まれる複数の小ブロックに対してシークンシャルに各ブロック内のデータの音声合成処理を順次行う。1つの無声音素片の読み出しは小ブロック内に格納される音声素片の長さ、くり返し回数、のデータ(データ  $\#1$  - データ  $\#m$ ) を基にして行う。

まず、アドレスカウンタ5にデータ  $\#1$  のアドレスが設定され、音声素片の長さ(データの数)に達するまで順次アドレスカウンタをカウントアップしてデータ  $\#m$  まで順方向に読み出すと、次はアドレスカウンタを順次カウントダウンしてデータ  $\#1$  に達するまで逆方向に読み出す。

すなわち、このようにして分析区間  $T$  の長さ(例えば30 ms)の無声音素片波形に対応するデータを読み出す。これを前記くり返し回数だけくり返すことにより、1つの小ブロックに対応する無声音素片の読み出しを行う。これを1つの音声ファイルに対応する大ブロック内の複数の小ブロックに対してシークンシャルに読み出し動作を行

い、有声音素片、無声音素片のデータを読み出すことにより、1つの単語等の音声ファイルの読み出し処理が終了する。このように、記憶装置より出力される各データはそのデータの形式に応じ、直接、又は復り器(図示せず)を介してD/A変換器6に入力され、更にローパスフィルタ7、音声出力部8により合成された音声として出力される。

以上説明したように、本発明によれば、無声音の音声素片を前記の方法によって波形対称化することによって合成音声の品質は劣化せず、また波形を対称にしたためその波形を記憶するための音声合成装置の記憶装置は記憶領域が半分になる利点がある。

本発明は、合成音声の品質を劣化させずに音声素片の記憶領域を軽減することができるので、すべての音声素片型の音声合成装置に使用することができ。

#### 4. 図面の簡単な説明

第1図は文系単位で記憶装置に言葉を格納したときの説明図、第2図は音節単位で記憶装置に音

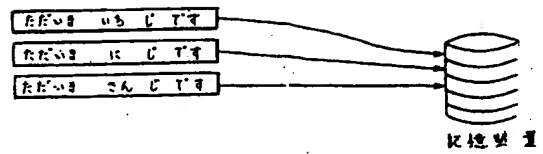
素を格納したときの説明図、第3図は文章、単語、音節単位で記憶装置に言葉を格納したときの説明図、第4図は音声波形例を示した図、第5図(a)は有声音波形例を示した図、第5図(b)は無声音波形例を示した図、第6図は記憶装置に音声を格納するときのブロック図、第7図(a)は無声音波形 f(t) を示した図、第7図(b)は無声音 f(t) の時間軸を逆にした時の波形 f(-t) を示した図、第7図(c)は無声音対称波形 g(t) を示した図、第8図は本発明の音声合成装置の記憶装置の記憶構造を示した図、第9図は本発明の音声合成装置のブロック図。

特許出願人 沖電気工業株式会社

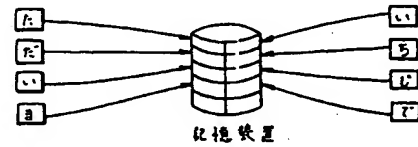
代理人 鈴木 敏 明



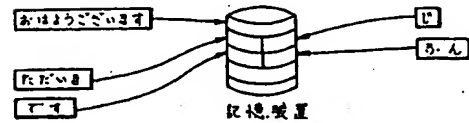
第1図



第2図



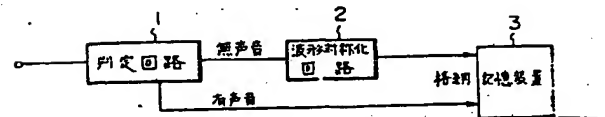
第3図



第4図



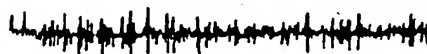
第6図



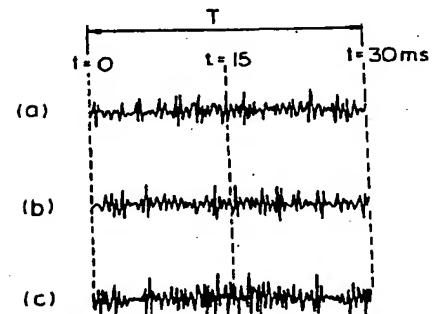
第5図(a)



第5図(b)



第7図



特開昭57-163299 (5)

手続補正書(自発)

昭和 56.7.7 日

特許庁長官 殿

1. 事件の表示  
昭和56年 特 許 願第 047143 号

2. 発明の名称  
音声合成装置

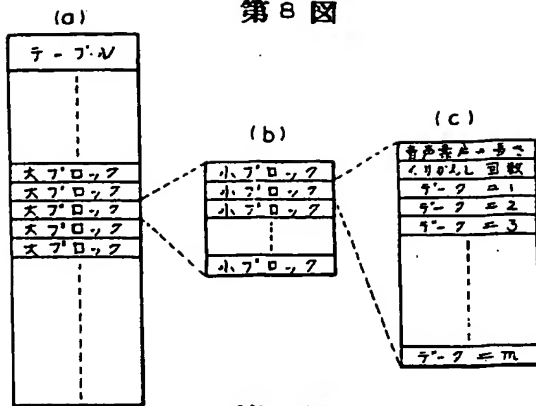
3. 補正をする者  
事件との関係 特 許 出 願 人  
住 所 (〒105) 東京都港区虎ノ門1丁目7番12号  
名 称 (029) 沖電気工業株式会社  
代表者 取締役社長 三宅正男

4. 代 理 人  
居 所 (〒105) 東京都港区虎ノ門1丁目7番12号  
氏 名 (6892) 弁理士 鈴木 敬 明  
電話 501-3111(大代表)

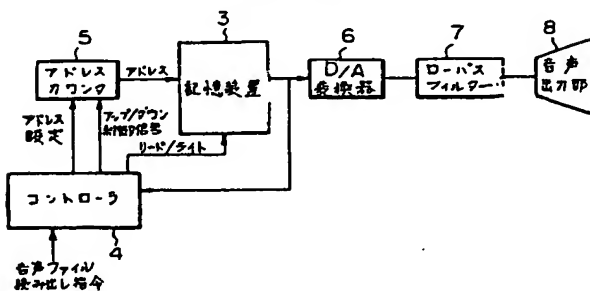
5. 補正の対象  
明細書中「発明の詳細な説明」および「図面の簡単な説明」の各欄

6. 補正の内容  
1) 明細書第9頁第6行に、「回数、の」とあるのを  
「回数、音声素片の」と補正する。  
2) 同書第11頁第8行に「/(-1)」とあるのを  
「/(30-1)」と補正する。

第8図



第9図



**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

Japanese Laid Open Publication (Kokai) No. S57-163299

Title of the Invention : Voice Synthesizing Apparatus

What is claimed : A voice synthesizing apparatus which outputs a synthetic sound by memorizing a large number of units which can reproduce pieces of silent phoneme and by selecting a piece of a silent phoneme corresponding to the voice which should be synthesized, wherein each said unit expresses  $1/2$  the time length of the arithmetical average of a piece of a natural silent phoneme and of a reversed time-axis of the corresponding piece of a natural silent phoneme, further comprising;

a means to reproduce the piece of a silent phoneme by reading a series of data of a desired unit in a direction of order and in an opposite direction, once in each direction in a sequential manner.

**THIS PAGE BLANK (USPTO)**